A Proof of Huygens' Principle

Liu Zhenxiu
Department of Mathematics, Hunan Normal University,
Changsha 410006, China

Fresnel and Kirchhoff made the proofs for Huygens' principle for the propagation of light respectively. In this paper, another proof for this principle is given, in which, by using Poisson's formula for the three-dimensional wave equation with initial-value conditions, the back-travelling waves are cancelled by the conterterms.

PACS. 4210-Optical propagation and transmission in homogeneous media

1. Introduction

According to Huygens' principle for the propagation of light, every point illuminated by an incident primary wavefront becomes the source of a secondary wavelet such that the primary wavefront at a later time results from the superposition of amplitudes of all wavelets. It is generally agreed that back-travelling waves would appear from the above given principle. To disprove this, Fresnel and Kirchhoff made a proof for the principle[1, 2]. Now we try to give another proof for the principle that when a point light source at P_o, emits a series of spherical wavelets, the effect at a point P could be found in either of two ways, First allow one of these wavefronts to proceed until it reaches P. Second, divide one wavefront into small areas, so that each area can generate a wavelet. Each of these wavelets then produces effects at P which can be summed, according to the principle of superposition, to give the total effect of the wavelets at P. These two methods must yield identical results. For this purpose, we should state Huygens' principle as follows.

Suppose that the initial-value problem

$Lu(x, y, z, t) = 0 \quad t > 0, \quad -\infty < x, y, z < \infty$

$u(x, y, z, 0) = \varphi(x, y, z,) \quad (x, y, z) \in D$

$u_t(x, y, z, 0) = \psi(x, y, z,) \quad (x, y, z) \in D$

where L is an operator and D is a domain in R^3, has the unique solution

$u(x, y, z, t) = f(x, y, z, t) \quad (x, y, z) \in D_{t.}$

Here D_t is the domain in R^3 at time t which is enclosed between the front and rear edges of the wave. Taking $t_1 > 0$ as the initial instant of time, then the unique solution of the following initial-value problem

$$Lu(x,y,z,t) = 0, \quad t > t_1, \quad -\infty < x,y,z < \infty$$
$$u(x,y,z,t_1) = f(x,y,z,t_1) \quad (x,y,z) \in D_{t_1}$$
$$u_t(x,y,z,t_1) = f_t(x,y,z,t_1) \quad (x,y,z) \in D_{t_1}$$

is also $f(x,y,z,t) \quad (x,y,z) \in D_t$. In particular, if we take $t = t_2 (> t_1)$, then

$$u(x,y,z,t_2) = f(x,y,z,t_2) \quad (x,y,z) \in D_{t_2}$$

That is to say, any point on a wave surface at t_1 behaves like a new source from which secondary waves spread out in a way satisfying the original wave equation.

To make this obvious, we first analyse the one-dimensional wave equation with initial-value conditions.

2. Initial-value problem of the one-dimensional wave equation

Consider the initial-value problem of the one-dimensional wave equation

$$\begin{aligned} &u_{tt} = a^2 u_{xx} \quad t > 0, \quad -\infty < x < \infty \\ &u(x,0) = \varphi(x) \quad -\infty < x < \infty \\ &u_t(x,0) = \psi(x) \quad -\infty < x < \infty. \end{aligned} \tag{1}$$

The solution of the problem, by D'Alembert's formula, is

$$u(x,t) = \frac{1}{2}[\varphi(x+at) + \varphi(x-at)] + \frac{1}{2a}\int_{x-at}^{x+at} \psi(\xi)d\xi \tag{2}$$

To simplify the problem, let ψ(x)=0. Then, at t=t_1, we have

$$u(x,t_1) = \frac{1}{2}[\varphi(x+at_1) + \varphi(x-at_1)]. \tag{3}$$

The function $\varphi(x-at_1)$ is a wave moving to the right with a speed a and the function $\varphi(x+at_1)$ is a wave moving to the left with a speed a.

The initial perturbation φ(x) for $t_1 > 0$ splits, so to speak, into two similar perturbations $(\frac{1}{2})\varphi(x \pm at_1)$, each with half of the intensity.(see Fig.1,2)

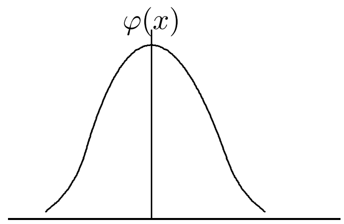


Fig. 1

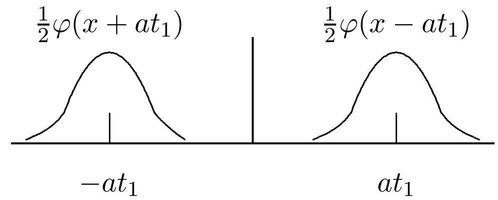


Fig. 2

Consider the case when $t = t_2 > t_1$ again. On the one hand, by formula (2), we have

$$u(x,t_2) = \frac{1}{2}[\varphi(x+at_2) + \varphi(x-at_2)] \tag{4}$$

On the other hand, we may take $u(x,t_1)$ as initial perturbation, and form the initial-value problem

$$\begin{aligned} &u_{tt} = a^2 u_{xx} \quad t > t_1, \quad -\infty < x < \infty \\ &u(x,t_1) = f(x,t_1) \quad -\infty < x < \infty \\ &u_t(x,t_1) = f_t(x,t_1) \quad -\infty < x < \infty, \end{aligned} \tag{5}$$

where $f(x,t_1) = \frac{1}{2}[\varphi(x+at_1) + \varphi(x-at_1)]$.

We transform the initial-value problem (5) to an initial-value problem in which the initial time is $\tau = 0$. To this end we will introduce a new variable $\tau = t - t_1$ instead of the variable t. Under this change of variable, the initial-value problem (5) for the function

$$\tilde{u}(x,\tau) = u(x,\tau + t_1)$$

becomes

$$\begin{aligned} &\tilde{u}_{\tau\tau} = a^2 \tilde{u}_{xx} \qquad \tau > 0, \quad -\infty < x < \infty \\ &\tilde{u}(x,0) = f(x,t_1) \quad -\infty < x < \infty \\ &\tilde{u}_\tau(x,0) = f_t(x,t_1) \quad -\infty < x < \infty \end{aligned} \tag{6}$$

Using D'Alembert's formula again, we have

$$\tilde{u}(x,\tau) = \frac{1}{2}[f(x+a\tau,t_1) + f(x-a\tau,t_1)] + \frac{1}{2a}\int_{x-a\tau}^{x+a\tau} f_t(\xi,t_1)d\xi$$

Let t = t_2. Then

$$\begin{aligned}u(x,t_2) &= \tilde{u}(x,t_2-t_1)\\ &= \frac{1}{2}f[x+a(t_2-t_1),t_1]+\frac{1}{2}f[x-a(t_2-t_1),t_1]\\ &+\frac{1}{2a}\int_{x-a(t_2-t_1)}^{x+a(t_2-t_1)} f_t(\xi,t_1)d\xi\end{aligned} \tag{7}$$

Substituting $f(x,t_1)=\frac{1}{2}[\varphi(x+at_1)+\varphi(x-at_1)]$, we find that

$$\begin{aligned}u(x,t_2) = &\frac{1}{4}\varphi(x-at_2)+\frac{1}{4}\varphi(x-2at_1+at_2)+\frac{1}{4}\varphi(x+2at_1-at_2)\\ &+\frac{1}{4}\varphi(x+at_2)-\frac{1}{4}\varphi(x-2at_1+at_2)+\frac{1}{4}\varphi(x+at_2)\\ &+\frac{1}{4}\varphi(x-at_2)-\frac{1}{4}\varphi(x+2at_1-at_2)\end{aligned} \tag{8}$$

We can easily see that the initial perturbations $f(x,t_1)=(1/2)[\varphi(x+at_1)+\varphi(x-at_1)]$ for $t_2>t_1$ split into eight perturbations. The first four terms rise from the initial value $f(x,t_1)$, (as shown in Fig.3.a). The last four terms rise from the initial speed $f_t(x,t_1)$.(Fig.3.b)

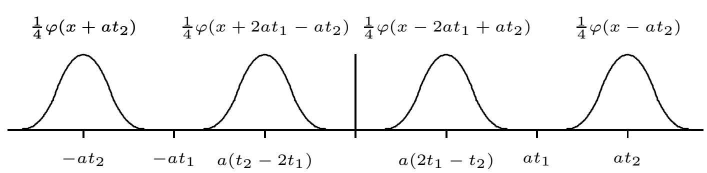


Fig. 3.a

Formula (8), and Fig.3.a , and Fig.3.b show there are back-traveling waves in the secondary wavelets, i.e. $\frac{1}{4}\varphi(x+2at_1-at_2)$ and $\frac{1}{4}\varphi(x-2at_1+at_2)$,rising from the initial value f(x,t_1). However, there are also two counterterms, $-\frac{1}{4}\varphi(x-2at_1+at_2)$ and $-\frac{1}{4}\varphi(x+2at_1-at_2)$, rising from the initial speed f_t(x,t_1). So the back-traveling waves are cancelled by each other. The final result is

$$u(x,t_2)=\frac{1}{2}[\varphi(x+at_2)+\varphi(x-at_2)] \tag{9}$$

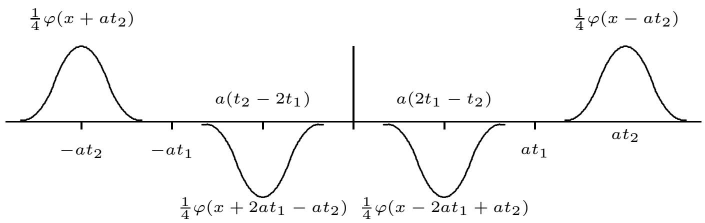


Fig. 3.b

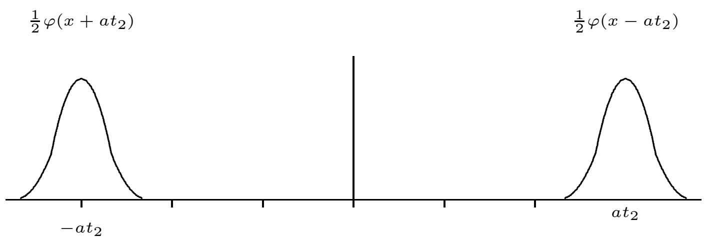


Fig. 3.c

which agrees with formula (4).(as shown in Fig.3.c)

The proof is similar when $\psi(x) \neq 0$.

3. Proof of Huygens' principle

Let P_0(Fig.4) be a point light source at which spherical monochromatic scalar waves are being generated continually beginning at t=0. At $t = t_1 > 0$, the sphere of radius ct_1, with centre at P_0, and called $S_{ct_1}^{P_0}$, is full of light, where c is the velocity of light. The light disturbance at a point Q within the sphere or on the spherical surface may be represented by

$$f(r,t_1) = \frac{A\sin(\omega t_1 - kr)}{r},$$

where A is the amplitude at unit distance from the source, and $r = |P_0Q|$. As usual, ω is the angular frequency and $k = \omega/c$.

We shall determine the light disturbance at a point P at $t_2 > t_1$. On the one hand, by allowing the light to proceed until it reaches P, the disturbance must be

$$f(R,t_2) = \frac{A\sin(\omega t_2 - kR)}{R} \tag{10}$$

where $R = |P_0P|$.

On the other hand, the "light sphere"$S^{P_0}_{ct_1}$ may be regarded as the initial disturbance and the scalar wave satisfies the wave equation $\Box f = 0$, where

$$\Box = \frac{\partial^2}{\partial t^2} - c^2(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2})$$

is the D'Alembert's operator. Therefore we form the initial-value problem

$$\begin{aligned} &\Box u(x,y,z,t) = 0 \quad t > t_1, \quad -\infty < x,y,z < \infty \\ &u(x,y,z,t_1) = f(r,t_1) \quad (x,y,z) \in S^{P_0}_{ct_1} \\ &u_t(x,y,z,t_1) = f_t(r,t_1) \quad (x,y,z) \in S^{P_0}_{ct_1} \end{aligned} \tag{11}$$

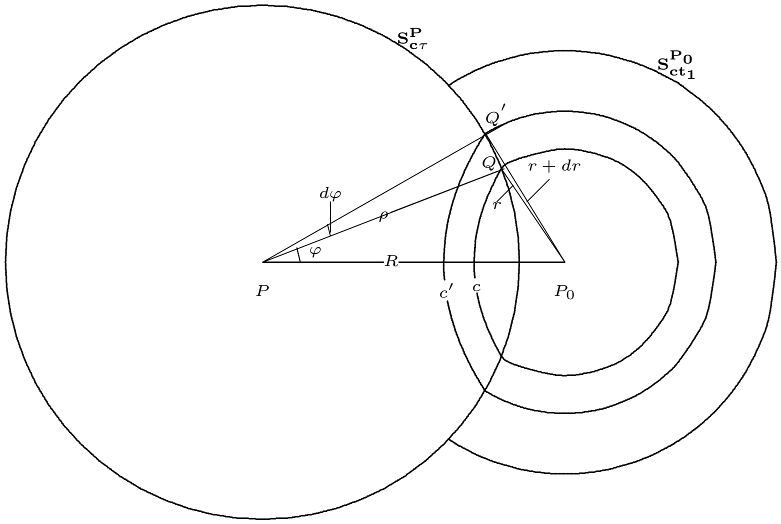
Let $\tau = t - t_1$ and $u(x,y,z,t) = u(x,y,z,\tau + t_1) = \tilde{u}(x,y,z,\tau)$. Then the above initial-value problem becomes

$$\begin{aligned} &\Box \tilde{u}(x,y,z,\tau) = 0 \quad \tau > 0, \quad -\infty < x,y,z < \infty \\ &\tilde{u}(x,y,z,0) = f(r,t_1) \quad (x,y,z) \in S^{P_0}_{ct_1} \\ &u_\tau(x,y,z,0) = f_t(r,t_1) \quad (x,y,z) \in S^{P_0}_{ct_1} \end{aligned} \tag{12}$$

According to Poisson's formula for the three-dimensional wave equation with initial-value conditions[3,4], the above initial-value problem has the solution

$$\tilde{u}(p,\tau) = \frac{1}{4\pi c}[\frac{\partial}{\partial \tau}\iint\limits_{S^P_{c\tau}} \frac{1}{\rho} f(r,t_1)ds + \iint\limits_{S^P_{c\tau}} \frac{1}{\rho} f_t(r,t_1)ds] \tag{13}$$

where $S^p_{c\tau}$ is a spherical surface, centered on P, with a radius ρ=cτ.(Fig. 4)



To evaluate the two integrals on the right-hand side of formula (13), we construct a sphere of radius $r(r < ct_1)$ with center P_0 . This sphere cuts the sphere $S^P_{c\tau}$ in a circle c centered on P_0P, and passing through a point Q on the spherical surface $S^P_{c\tau}$. Let the light disturbance at the arbitrary point on the circle c be represented by $f(r, t_1) = Asin(\omega t_1 - kr)/r$. Another circle c' can be constructed using a sphere, centered on $P_{0,}$ of radius r+dr. This circle is also centered on P_0P and passes through a point Q' on the $S^P_{c\tau}$. There is a ring zone between the c and the c' on the $S^P_{c\tau}$. Let the area of the ring zone be ds, and let φ be the angle between P_0P and PQ. From Fig.4

$$r^2 = \rho^2 + R^2 - 2\rho Rcos\varphi.$$

So that

$$rdr = \rho Rsin\varphi d\varphi$$

and therefore

$$ds = 2\pi\rho\sin\varphi d\varphi = \frac{2\pi\rho r}{R}dr. \tag{14}$$

Now we must discuss two cases.

Case I $R + \rho \leq ct_1$.

Then the two integrals in right-hand side of formula (13) are taken over the whole spherical surface $S^p_{c\tau}$ and r varies from $R-c\tau$ to $R+c\tau$. It follows that

$$\begin{aligned}
\tilde{u}(p,\tau) &= \frac{1}{4\pi c}[\frac{\partial}{\partial\tau}\int_{R-c\tau}^{R+c\tau}\frac{1}{\rho}\cdot\frac{A}{r}\sin(\omega t_1-kr)2\pi\frac{\rho r}{R}dr \\
&+ \int_{R-c\tau}^{R+c\tau}\frac{1}{\rho}\cdot\frac{A\omega}{r}\cos(\omega t_1-kr)2\pi\frac{\rho r}{R}dr] \\
&= \frac{A}{2cR}[\frac{\partial}{\partial\tau}\int_{R-c\tau}^{R+c\tau}\sin(\omega t_1-kr)dr+\int_{R-c\tau}^{R+c\tau}\omega\cos(\omega t_1-kr)dr] \\
&= \frac{A}{2R}[\sin(\omega t_1-kR-kc\tau)+\sin(\omega t_1-kR+kc\tau) \\
&- \sin(\omega t_1-kR-kc\tau)+\sin(\omega t_1-kR+kc\tau)].
\end{aligned}$$

Substituting $kc=\omega$ and let $t=t_2$, we find that

$$\begin{aligned}
u(p,t_2) &= \frac{A}{2R}[\sin(2\omega t_1-\omega t_2-kR)+\sin(\omega t_2-kR) \\
&- \sin(2\omega t_1-\omega t_2-kR)+\sin(\omega t_2-kR)] \\
&= \frac{A}{R}\sin(\omega t_2-kR).
\end{aligned} \tag{15}$$

Case II $R+\rho>ct_1$.

Then the two integrals in the right-hand side of formula (13) are taken over a portion of the spherical surface $S^P_{c\tau}$and r varies from $R-c\tau$ to $R+c\tau-\gamma$, where $0<\gamma<2c\tau$ and $R+c\tau-\gamma=ct_1$. It follows that

$$\begin{aligned}
\tilde{u}(p,\tau) &= \frac{A}{2cR}[\frac{\partial}{\partial\tau}\int_{R-c\tau}^{R+c\tau-\gamma}\sin(\omega t_1-kr)dr \\
&+ \int_{R-c\tau}^{R+c\tau-\gamma}\omega\cos(\omega t_1-kr)dr] \\
&= \frac{A}{2R}[\sin(\omega t_1-kR-kc\tau+k\gamma)+\sin(\omega t_1-kR+kc\tau) \\
&- \sin(\omega t_1-kR-kc\tau+k\gamma)+\sin(\omega t_1-kR+kc\tau)].
\end{aligned}$$

Substituting $kc = \omega$ and let $t = t_2$, we find that

$$\begin{aligned} u(p, t_2) &= \frac{A}{2R}[\sin(2\omega t_1 - \omega t_2 - kR + k\gamma) + \sin(\omega t_2 - kR) \\ &\quad - \sin(2\omega t_1 - \omega t_2 - kR + k\gamma) + \sin(\omega t_2 - kR)] \\ &= \frac{A}{R}\sin(\omega t_2 - kR). \end{aligned} \tag{16}$$

Both formulae (15) and (16) agree with the formula (10).

We can rewrite the two terms in the right-hand side of the first equal-sign in formula (15). Thus

$$\frac{A}{2R}\sin(2\omega t_1 - \omega t_2 - kR)$$

and

$$-\frac{A}{2R}\sin(2\omega t_1 - \omega t_2 - kR)$$

become

$$-\frac{A}{2R}\sin k[R + c(t_2 - 2t_1)]$$

and

$$\frac{A}{2R}\sin k[R + c(t_2 - 2t_1)]$$

respectively. We can see that the terms are the back-traveling waves and they cancel each other. The case is similar in the formula (16).

It should be note that, more generally if we let the primary wave be

$$\frac{f(r - ct)}{r}.$$

where f is an arbitrary function, we can also obtain the same result.

E-mail: zxliu2002@hotmail.com

References

1. Max Born and Emil Wolf, Principle of Optics, 7th (expanded) edition, Cambridge University Press, 1999, 413-422.

2. Eugene Hecht Alfred Zajac, Optics, 1st ed. (Addison-Wesley Publishing Company 1976), Chap. 10.

3. V. S. Vladimirov, Equations of Mathematical Physics, Translated from the Russian by Engene Yankovsky, Mir Publishers Moscow, 1984, p.202, formula(13, 19).

4. Morse and Feshhach, Method of Theoretical Physics, McGraw-Hill, 1953, p.847, formula(7, 3, 20).